International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-026-03781-2

ORIGINAL ARTICLE

Exploring performance and trust in autonomous surgical robotic systems: the role of visuo-haptic guidance

Frank Papenmeier[2] · Katharina Hagmann[1] · Lisa Blossey[2] · Adrian S. Bauer[1] · Andreea E. Potinteu[2,3] · Alexander Kirst[1] · Daniel Leidner[1] · Markus Huff[2,3] · Julian Klodmann[1]

Received: 10 January 2026 / Accepted: 5 August 2026
© The Author(s) 2026

Abstract

Purpose: Incorporating task-level autonomy into teleoperation systems for robot assisted laparoscopic surgery requires operators to effectively manage transitions from task-level autonomy to teleoperation control, particularly during system failures. This study investigated the impact of the level of guidance during take-over situations on participants' performance and their trust in the teleoperation system.

Methods: $N = 34$ students, unfamiliar with the system and task, performed a ring transfer task under two conditions: visual guidance, where they monitored the system visually, and visuo-haptic guidance, where they also received haptic feedback by holding the input device as the system operated in task-level autonomy. During some trials, system failures occurred, requiring participants to take over control.

Results: Results showed that participants noticed system failures faster with visuo-haptic guidance compared to visual guidance. Although visuo-haptic guidance did not improve task completion times or accuracy during task completion, it led to significantly higher trust in the teleoperation system and a stronger sense of perceived control. However, participants also reported higher physical demand and effort under visuo-haptic guidance, indicating an increased workload in this condition.

Conclusion: In conclusion, these findings reveal important trade-offs of the influence of haptic guidance during human oversight of teleoperated autonomous systems.

Keywords Teleoperation · Visuo-haptic guidance · Human-robot interaction · Trust in automation · System failures

Introduction

In recent years, the usage of robotically assisted surgical systems in laparoscopy (RASS-L) has expanded enormously. These telemanipulation systems serve as the surgeon's eyes and hands within the patient's body. Current research focuses on advancing the *Levels of Autonomy* (LoA) in these robotic systems. LoA categorize robotic systems from no autonomy (L0) through various levels, such as robot assistance (L1), task-level autonomy (L2), supervised autonomy (L3) and high level autonomy (L4) progressing toward full autonomy (L5) [1, 2]. This work focuses on LoA up to L2:

- L0: The surgeon telemanipulates the instrument tip freely within its inherent workspace.
- L1: The robotic system provides shared control based on haptic augmentation.
- L2: The robotic system controls all *Degrees of Freedom* (DoF) to execute a predefined task.

Regulatory standards, currently developed for *Artificial Intelligence* (AI) systems, emphasise *Human Oversight* as crucial requirement starting from L2. Particularly, the AI

Frank Papenmeier and Katharina Hagmann have equally contributed to this work.

✉ Frank Papenmeier
frank.papenmeier@uni-tuebingen.de

✉ Julian Klodmann
julian.klodmann@dlr.de

1 Institute of Robotics and Mechatronics, German Aerospace Center, Muenchener Str. 20, 82234 Wessling, Bavaria, Germany

2 Department of Psychology, University of Tuebingen, Schleichstr. 4, 72076 Tuebingen, Baden-Wuerttemberg, Germany

3 Leibniz-Institut für Wissensmedien, Schleichstr. 6, 72076 Tuebingen, Baden-Wuerttemberg, Germany

Published online: 07 November 2026

Act's Articles 3 and 6 of the European Union classify RASS-L high-risk AI systems and Article 14 defines basic requirements of Human Oversight [3]. Teleoperation systems can implement Human Oversight either *visually*, by observing the robotic system performing the task, or *visuo-haptically*, by both watching and being haptically guided as the robot executes the task. The latter requires haptic augmentation capabilities at the surgeon console such that Virtual Fixture (VF) based methods can be applied [4, 5].

Furthermore, the AI-Act's Article 14 states that surgeons must have the ability to decide to take over full control of the robotic system at any moment. Besides the surgeons, also the system itself may trigger transitions from L2 to L1 or even L0 in case of unforeseen events, such as system failures or sudden increase of uncertainty, prompting the surgeon to take over control. To do so effectively, the surgeon must quickly assess the situation and understand the current state of the surgery. This raises the question of whether the type of human oversight during task-level autonomy (L2), called *level of guidance* in the following, which can be either visual or visuo-haptic, impacts the surgeon's performance during takeover events.

Visual and visuo-haptic guidance can support robot assisted surgery training [5]. Addionally, adding haptic cues to shared control teleoperation can improve users' situational awareness and satisfaction [6]. Regarding takeover events, we draw on research on autonomous driving. This line of research distinguishes between two critical phases: the initial time required for takeover [7] and the time required for stabilizing control after the driver is back in control [8]. The initial takeover time is increased when drivers perform non-driving tasks during autonomy and decreased when takeover requests are provided [7]. Once drivers are back in control they take some time to stabilize their lateral control of the vehicle [8], showing the importance of measuring not only the time required for takeover but also to investigate the performance after users are back in control.

Applying these insights to surgical robotics, we propose that visuo-haptic guidance in L2 implemented at the input device of the RASS-L may serve as an implicit takeover request due to the sudden stop of guidance force during system failure, which might accelerate its recognition. In addition, the continuous visuo-haptic coupling prior to failure might help to maintain the situational awareness needed for improved post-takeover control and performance.

Beyond these objective performance metrics, the effective adoption of autonomous systems relies fundamentally on user trust. People often assume that autonomous systems novel to them are perfect, thereby showing high initial trust [9]. Further, research indicates that trust levels are heavily contingent on system performance [10] and the frequency of system failures [11, 12]. This study investigated whether system failures during L2 that require user takeover result in a loss of trust in the teleoperation system and whether this change in trust is influenced by the type of guidance experienced during L2. Furthermore, the specific level of guidance provided during automation is likely to influence the user's perceived control. While purely visual guidance places the operator in a passive supervisory role, potentially creating a sense of detachment, visuo-haptic guidance maintains a physical coupling between the surgeon's hands and the robotic instruments. This study investigated whether continuous physical engagement results in higher subjective ratings of perceived control compared to visual guidance alone.

Present study and hypotheses

We designed an experiment using a hardware-in-the-loop simulator of the RASS-L research platform DLR MiroSurge to investigate user performance during takeover situations from L2 to L1 under two levels of guidance: visual and visuo-haptic (see Fig. 1) [13, 14]. Furthermore, the influence of system failures on trust and perceived control was investigated.

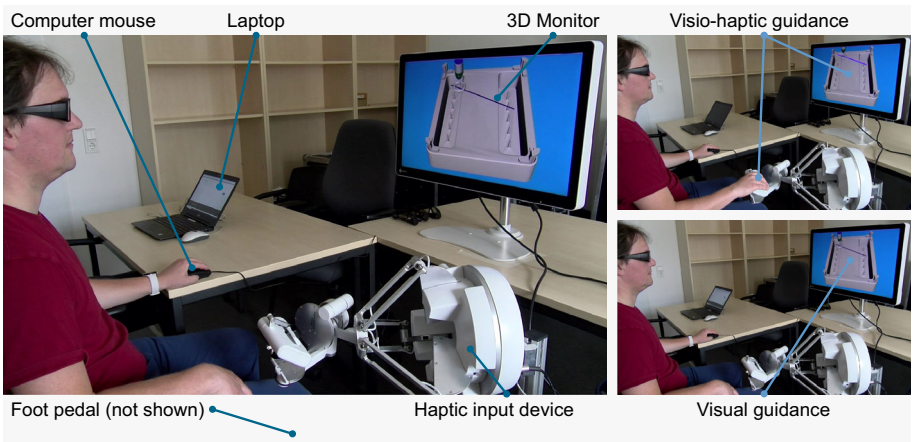


Fig. 1 Experimental setup of the simulator (left) described in further detail in Sec. Methods. Participants performed a ring transfer task using visual (bottom right) or visuo-haptic (top right) guidance

Regarding the influence of the level of guidance on the performance during takeover events, we predicted that participants notice system failure faster with visuo-haptic guidance (Hypothesis 1). Further, we predicted that visuo-haptic guidance leads to faster task completion (Hypothesis 2) and less path deviation during task completion (Hypothesis 3) in L1 shared control teleoperation after takeover.

Because a reduction in technological novelty reduces initial trust [9], we predicted that higher prior experience with autonomous systems is associated with lower initial trust in autonomous systems in general, thereby resulting in a reduced loss of trust following a system failure (Hypothesis 4). Furthermore, we predicted that trust in the teleoperation system decreases after system failures were experienced (Hypothesis 5a) and that a statistical interaction exists where this loss of trust is significantly higher for visual than for visuo-haptic guidance (Hypothesis 5b). Regarding perceived control, we predicted that visuo-haptic guidance leads to higher perceived control (Hypothesis 6).

Methods

This section provides a detailed description of the study design, which was preregistered at the Open Science Framework (https://osf.io/b3wvx).

Participants

A total of 39 participants were recruited from the University of Tübingen. We recruited naïve participants with no expertise in robot-assisted laparoscopy. To ensure participants could physically operate the right-handed haptic input device and properly view the stereoscopic 3D monitor layout, they were required to be right-handed and to have stereoscopic vision, which was assessed using the Randot SO-002 stereotest (Stereo Optical Co., Inc.). One participant failed this test and was therefore excluded from the study. Additionally, four participants were excluded from the analysis due to technical issues with the RASS-L simulator. Thus, the final sample consisted of 34 participants with an average age of 24.56 years ($SD = 6.08, range : 19 - 49$), with 24 identifying themselves as female, nine as male and one as non-binary. Participation in the experiment took approximately 75 min and participants received 15 € as compensation. Alternatively, psychology and cognitive science students could choose to receive course credits to fulfill the requirement to participate in empirical research during their degree program.

Robotic system

The experiment was conducted using a hardware-in-the-loop simulation (see Fig. 1) of a subset of the RASS-L research platform DLR MiroSurge [13, 14]. The subset consists of two kinematically simulated 7 DoF robotic arms, DLR MIRO, mounted to the side rail of an operating room table. One MIRO carries a stereoscopic endoscope to provide a 3D view of the patient's inside for the user. The second MIRO is equipped with a kinematically simulated drive unit DLR MICA actuating a surgical instrument, which integrates a 2 DoF wrist joint and a gripping DoF. The user controls the 6 DoF pose of the instrument's *Tool Center Point* (TCP) $\boldsymbol{x}$ via a Force Dimension sigma.7 haptic input device, which provides force and torque feedback in 7 DoF. A software layer integrates a position controller that maps the motion of the input device to the instrument TCP, using inverse kinematics. This inverse kinematics also preserves the motion constraint at the incision point and restores the hand-eye coordination. The virtual scene was rendered in 3D and presented via a 3D monitor.

During the experiment the *Autonomy and Shared Control Parameterization Engine* (ASCOPE) of the MiroSurge system was utilized which was introduced and described in detail in [4]. ASCOPE provides L2, L1 and L0 and allows to trigger continuous transitions between all three LoAs at any point in time during the experiment. To this end, ASCOPE incorporates *Virtual Fixtures* (VFs) to assist the user in the provided LoA. A VF implements a virtual *Spring Damper System* (SDS) acting between the instrument TCP $\boldsymbol{x}$ and a virtual proxy $\boldsymbol{x}_p$ [15]. If the SDS is deflected, the generated wrench (i.e. Cartesian forces and torques in 6 DoF) are projected to the handle of the haptic input device. As such, the user perceives the generated wrench haptically. Splines are used to model the required trajectories for completing the given tasks in L1 and L2.

In L0 the user telemanipuates the instrument TCP $\boldsymbol{x}$ freely and no VF is active. In L1 a guidance VF is parameterized, such that $\boldsymbol{x}_p$ is fixtured via an SDS on the spline in 5 DoF. The sixth DoF tangential to the spline is kept free and can therefore be controlled by the user (Fig. 2). Task-level autonomy (L2) is enabled by a virtual mass attached to $\boldsymbol{x}_p$, applying a virtual feed-forward force $\boldsymbol{f}_{ff}$ to it and activating the SDS in the tangential direction. The force causes $\boldsymbol{x}_p$ to move and thus deflects the SDS. The computed wrench of the SDS is projected to the haptic device such that its handle moves and thus the instrument TCP $\boldsymbol{x}$ moves accordingly (Fig. 2).

The latter enables visuo-haptic guidance by letting the user hold the handle of the input device while the task is executed (see Fig. 1 and supplementary video).

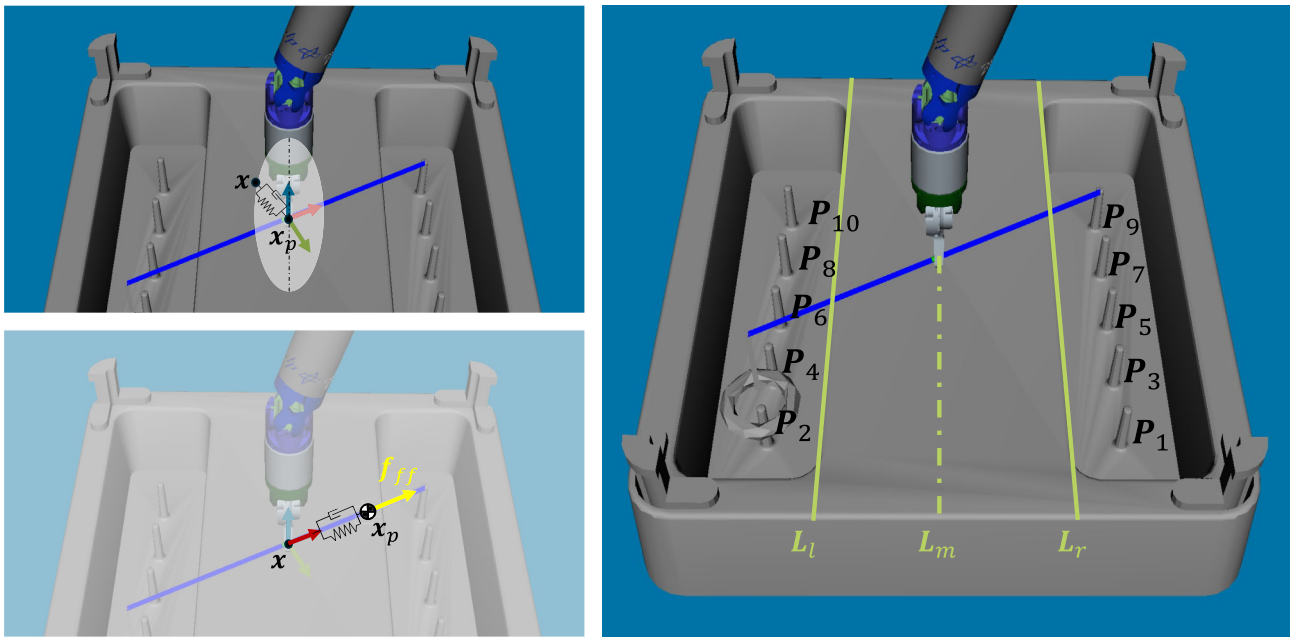


Fig. 2 Simplified schemes of the virtual fixtures in L1 shared control teleoperation (left, top) and in L2 task-level autonomy (left, bottom). Illustration of the ring transfer task (right): P_1 to P_{10} are the possible pegs where to pick or place the ring. L_l and L_r span an area around the middle line of the training pad L_m. In this area the system could stop and the user had to take over

Ring transfer task and levels of guidance

Participants used the robotic system to perform a ring transfer task based on the bandtwist training pad of the Luebeck Toolbox [16], as shown in Fig. 2 on the right. This standardized training task, adapted from other training curricula, such as the Fundamentals of Laparoscopy program from North America, has become established in the surgical robotics research community to assess the general suitability of novel methods to implement different levels of autonomy [17]. Furthermore, the task at hand not only allows conclusions for teleoperation systems in general, it also allows to initially involve non-expert users due to its simplicity.

The task consisted of multiple trials with each trial requiring to pick up a ring from one of the ten pegs ($P_1 - P_{10}$) and place it on a predefined peg on the opposite side. The system performed this task autonomously in L2, but in some trials the execution stopped, and without any further notification, the participant needed to take over control and complete the task in L1. During all trials, the gripper opened and closed autonomously while picking up or placing the ring.

Each trial started with the instrument and input device positioned in the starting position, and the ring at a random starting peg. The participant pressed the foot pedal to couple the input device and the instrument TCP, thereby starting the autonomous task execution. During transport, the system could stop task execution, and participants were instructed to indicate that they noticed the system stop by pressing a computer mouse button that was placed in their left hand. After the mouse click, they placed their hand on the input device (or the hand was already there in the visuo-haptic condition), then pressed the foot pedal to couple the systems again, and completed the transfer and placing task in L1. Between trials, the screen turned grey and the input device and instrument returned to the initial position and orientation. Participants removed their hand from the input device while the screen was grey (see Fig. 1 and supplementary video for further clarification of the workflow).

To address the research questions, we manipulated the level of guidance in a one-factorial within-participants design with two conditions: *visual guidance* and *visuo-haptic guidance*. In visual guidance, the system performed the ring transfer task autonomously and participants visually observed the autonomous task execution. The visuo-haptic condition was identical to the visual condition with the only difference that participants placed their hand on the input device for additional haptic guidance. Thus, the input device moved autonomously in both conditions.

Questionnaires

All questionnaires were embedded into an online SoSci survey and presented to the participants on a laptop (https://www.soscisurvey.de/en/index).

Trust

We measured trust using the performance trust scale of the MDMT trust questionnaire [18]. The performance trust scale consists of the reliable subscale (4 items) and the competent subscale (4 items). In our questionnaire, we presented the items from the reliable subscale and the competent subscale alternately. For each participant, the level of trust was calculated as the average of the mean reliable and the mean competent subscale scores. We used an 8-point scale ranging from 0 ("Not at all") to 7 ("Very much") with an additional "does not fit" option, with larger scores indicating a larger level of trust.

Perceived control

We measured perceived control with two items ("I felt like I had control over the system" and "I consider the system to be an assistant"). The 8-point scale ranged from 0 ("Not at all") to 7 ("Very much"). The mean value of both items served as the main dependent variable for perceived control.

Prior experience

We measured prior experience with autonomous systems in general with two items: "I have prior experience with autonomous systems" (0 = "None" to 7 = "A lot") and "I use autonomous systems" (0 = "Never" to 7 = "Very often"). The mean value of both items served as the main dependent variable for prior experience.

NASA-Task Load Index

We measured subjective workload using the NASA-Task Load Index (NASA-TLX) questionnaire [19]. Participants rated the six subscales (mental demand, physical demand, temporal demand, performance, effort, frustration) on a 21-point scale each, which corresponds to a rating of 0 to 100 because each tick divides the final scale (0 to 100) into increments of 5.

Procedure

Figure 3 illustrates the procedure of our study. After providing informed consent and completing a demographic questionnaire, participants filled in questionnaires on prior experience with and trust in autonomous systems in general (e.g., robotic vacuums, semi-autonomous driving, or virtual assistants) and performed the Randot stereo vision test. Next, they received written instructions on teleoperated RASS-L in general, the ring transfer task (Sect. Ring transfer task and levels of guidance), that the system will perform the task autonomously in general, but that the system might stop during the ring transport requiring participants to intervene as fast and precise as possible. Participants were instructed that they will interact with two systems, one

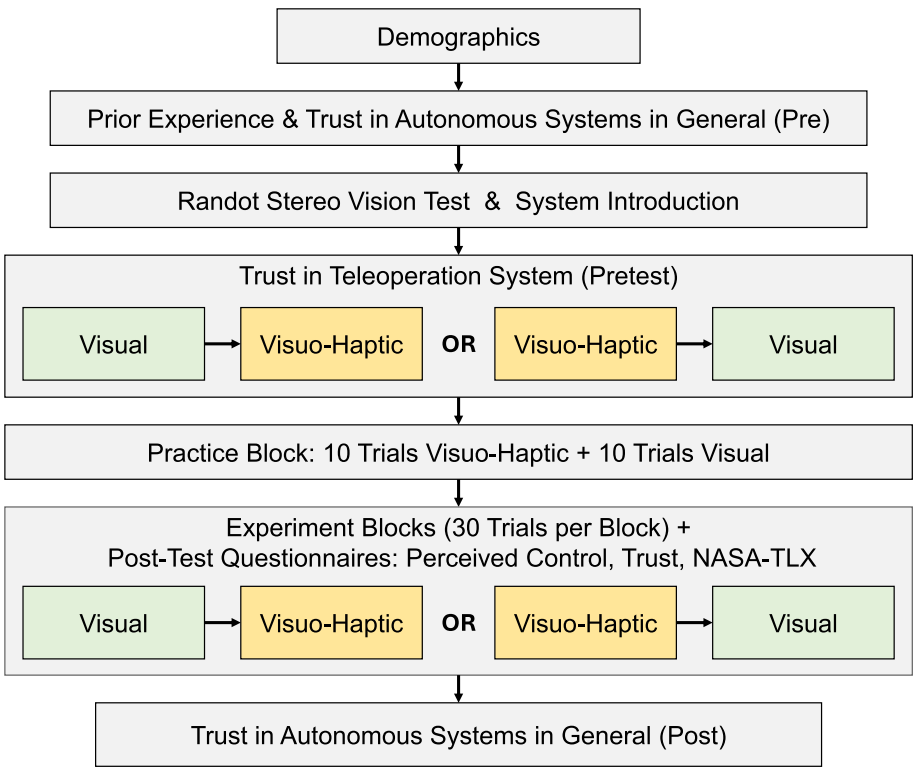


Fig. 3 Experimental procedure

called the visual system (where participants should visually supervise the system without touching the input device) and the other called the visuo-haptic system (where participants should visually supervise the system and touch the input device to be guided and feel how the system executes the task). Thereafter, participants filled in the pretests on trust for both systems in counterbalanced order.

Participants were seated at the system and received a brief oral introduction. To familiarize, they performed two practice blocks of 10 trials (including two system failure trials) each: first with visuo-haptic guidance (instructed to keep their wrist loose to allow the system to guide their hand), followed by visual guidance.

After a short break, participants performed the two experimental blocks (visual and visuo-haptic). The order of blocks was counterbalanced across participants. Each block comprised six sets of five trials. Trials for both blocks were generated before the experiment, following these rules: In each set, one experimental trial (i.e., trial containing a system failure requiring takeover) was paired with four filler trials. The order was randomized. For each trial, a start peg was randomly selected, with the destination being the opposite or a diagonal peg ($\pm$ 2). For example, if the ring was picked from P_5, it might be placed at P_6 (opposite), P_2 (-2), or P_{10} (+2). Experimental trials included the stop position, at which the system stopped while transitioning the ring due to a system failure, within a $\pm$ 2 cm corridor from the middle line (L_m) of the training pad. Figure 2 on the right shows the stop area, defined by L_l and L_r. Further, we balanced trials within each experimental block to ensure that the six experimental trials contained exactly one trial for each 3 (target: horizontal, +2, -2) x 2 (task direction: left-to-right, right-to-left) combination. Similarly the 24 filler trials were generated to contain four trials of each combination.

To ensure a fair within-participants comparison between the two guidance conditions, that is to ensure that takeover in the experimental trials was equally difficult for both blocks, the experimental trials in the second experimental block were mirrored versions of the first block (e.g., $P_1 \rightarrow P_6$ became $P_6 \rightarrow P_1$, with stop positions mirrored accordingly). Mirroring was chosen to reduce the risk of participants recognizing the experimental trials as system failure trials before the system failure occurred. Filler trials in the second block were new trials randomly generated following the same rules as in the first block.

After each block, participants were seated in front of the laptop and completed questionnaires on perceived control, trust, and NASA-TLX. Following the two experimental

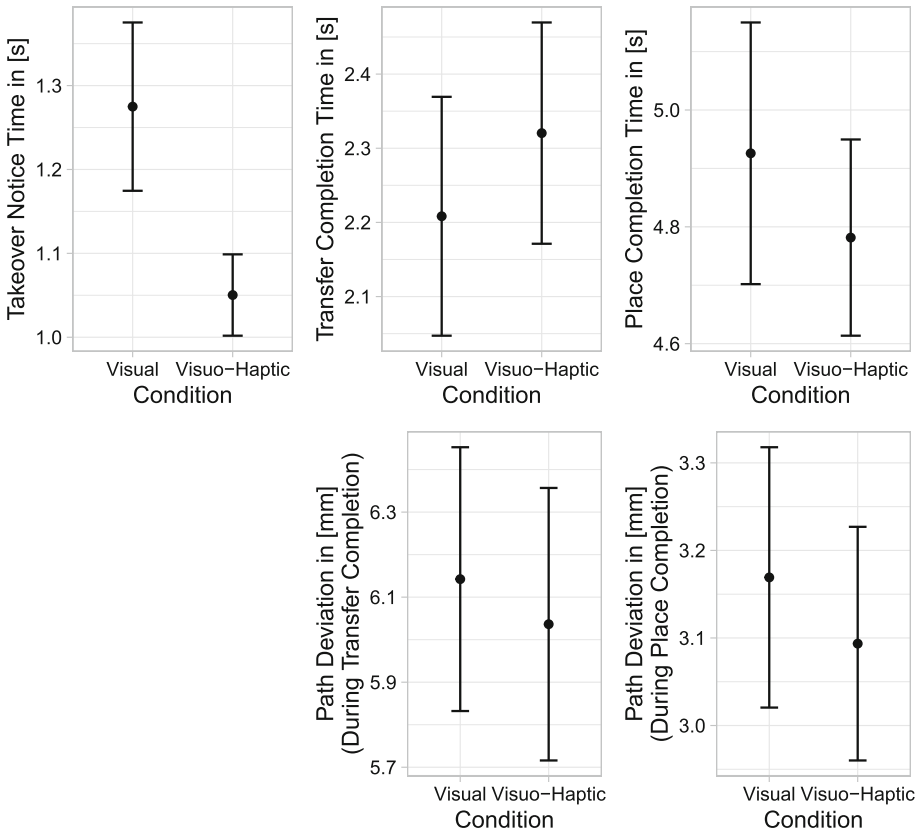


Fig. 4 Influence of guidance during L2 on the performance measures during takeover events and task completion in L1. Error bars indicate the standard error of the mean

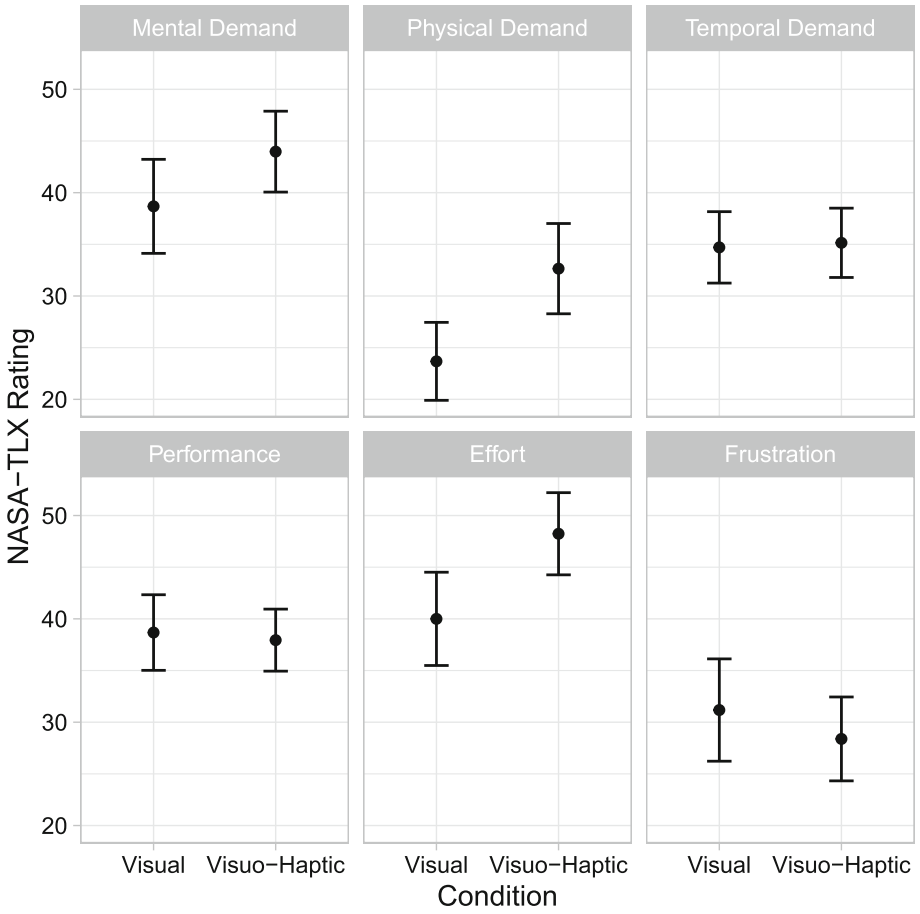


Fig. 5 Influence of guidance during L2 on the NASA-task load index ratings. Error bars indicate the standard error of the mean

Table 1 Paired t-test results comparing visual and visuo-haptic guidance across performance metrics and workload subscales

Category	Metric / Subscale	$t(33)$	p	d_z	95% CI
Performance	Takeover notice time	2.36	**.024**	0.40	[0.05, 0.74]
	Transfer completion time	−1.42	.165	−0.24	[−0.57, 0.10]
	Place completion time	1.26	.217	0.21	[−0.12, 0.54]
	Path deviation (transfer)	0.26	.799	0.04	[−0.29, 0.37]
	Path deviation (place)	0.72	.477	0.12	[−0.21, 0.45]
NASA-TLX	Mental demand	−2.01	.053	−0.34	[−0.67, 0.00]
	Physical demand	−3.04	**.005**	−0.51	[−0.86, −0.16]
	Temporal demand	−0.16	.878	−0.03	[−0.35, 0.30]
	Performance	0.24	.811	0.04	[−0.29, 0.37]
	Effort	−2.75	**.010**	−0.46	[−0.81, −0.11]
	Frustration	0.83	.412	0.14	[−0.19, 0.47]

blocks, participants once again filled in the questionnaire on trust in autonomous systems in general. Thereafter, they were debriefed about the goals of the experiment and given the choice to either retract their data or agree in the analysis of their data.

Results and discussion

Performance and workload

To evaluate the influence of the level of guidance during L2 (visuo-haptic vs. visual) on the objective performance metrics (see Fig. 4) and the subjective workload scales (see

Fig. 5), we conducted two-tailed paired t-tests for the respective dependent variables (see Table 1). All analyses were performed at a significance level of $\alpha = .05$. As effect size we report Cohen's d_z together with its 95% confidence interval.

Confirming Hypothesis 1, participants noticed the system failure faster with visuo-haptic guidance than with visual guidance. Contrary to Hypotheses 2 and 3, visuo-haptic guidance as compared to visual guidance during L2 did not significantly influence completion times and path deviations during transfer and placing in L1.

Regarding subjective workload, participants rated visuo-haptic guidance as significantly physically more demanding than visual guidance and reported significantly higher effort under the visuo-haptic condition. The effect on mental demand was marginal with a slightly higher mental demand for visuo-haptic guidance than visual guidance. No significant influence was found on temporal demand, performance, and frustration.

To summarize, visuo-haptic guidance was subjectively rated as more physically demanding and required more effort compared to visual guidance. However, it facilitated a faster detection of system failure, confirming the hypothesis that visuo-haptic feedback enhances the participant's ability to notice issues during task-level autonomy. Although an absolute reduction in detection time of approximately 200 ms may carry limited clinical significance during an isolated surgical task, from a human-factors perspective it provides a vital proof of concept, demonstrating that continuous physical coupling provides immediate tactile cues that successfully accelerate human cognitive processing during automation failures. Despite these differences, performance after the takeover was not significantly influenced by the type of guidance. Both guidance levels led to similar task completion times and path deviations, suggesting that once control is handed over, the level of guidance prior to the takeover event may not play a critical role in performance following the takeover event.

Trust

A linear mixed effect model was calculated for Hypothesis 4 with trust in autonomous systems in general as the dependent variable. The model included a fixed effect for time point (before vs. after experiencing system failure), prior experience with autonomous systems, and their interaction, with the participant's intercept as a random effect. Type II Wald chi-square tests evaluate the fixed effects. The results revealed a significant main effect of time point, $\chi^2(1) = 7.37$, $p = .007$, indicating that system failure had a significant impact on trust. The mean general trust was significantly lower before ($M = 5.28$, $SD = 0.72$) than after ($M = 5.67$, $SD = 0.76$) experiencing system failure. No significant main effect of prior experience, $\chi^2(1) = 3.20$, $p = .074$, was observed and the interaction between prior experience and time point was not significant, $\chi^2(1) = 0.15$, $p = .697$.

For Hypotheses 5a and 5b, a 2 (time point: before vs. after system failure; within) x 2 (level of guidance: visuo-haptic, visual; within) repeated-measures ANOVA was calculated with trust in the teleoperation system as the dependent variable. Following our pre-registered exclusion criteria, we removed one participant who provided more than two "does not fit" responses to the competent sub-scale of the MDMT for this analysis. Figure 6a illustrates the results. No significant main effect of time point was found, $F(1, 32) = 0.06$, $p = .806$, indicating no difference in trust before and after system failure. The interaction between time point and guidance was also non-significant, $F(1, 32) = 0.05$, $p = .821$. However, a significant main effect of guidance was found, $F(1, 32) = 19.37$, $p < .001$, with a significantly higher trust in the teleoperation system with visuo-haptic guidance ($M = 5.80$, $SD = 0.64$) compared to visual guidance ($M = 5.28$, $SD = 0.79$).

Trust in autonomous systems in general increased after participants read about and interacted with the robotic system, regardless of their prior experience. Higher trust in the teleoperation system was reported when visuo-haptic guidance was used compared to visual guidance. However, this increase in trust was not influenced by actual system interaction, as no significant change in trust was observed after participants physically engaged with the system. These findings suggest that the robotic system met user expectations set by the system description. Reading about the system was sufficient to boost trust, while actual interaction did not further alter this perception.

Perceived control

A two-tailed paired t-test was computed for Hypothesis 6 with perceived control as the dependent variable and guidance during interactive autonomy (visuo-haptic vs. visual) as the independent variable. Guidance had a significant effect on perceived control, $t(33) = -4.06$, $p < .001$, $d_z = -0.68$, 95% CI [−1.05, −0.31], with visuo-haptic guidance resulting in a significantly higher perceived control compared to visual guidance (see Fig. 6b).

Conclusion

Incorporating task-level autonomy into teleoperation systems for robot assisted laparoscopic surgery requires operators to effectively manage transitions from task-level autonomy (L2) to shared control teleoperation (L1), particularly during system failures. This study investigated the impact of

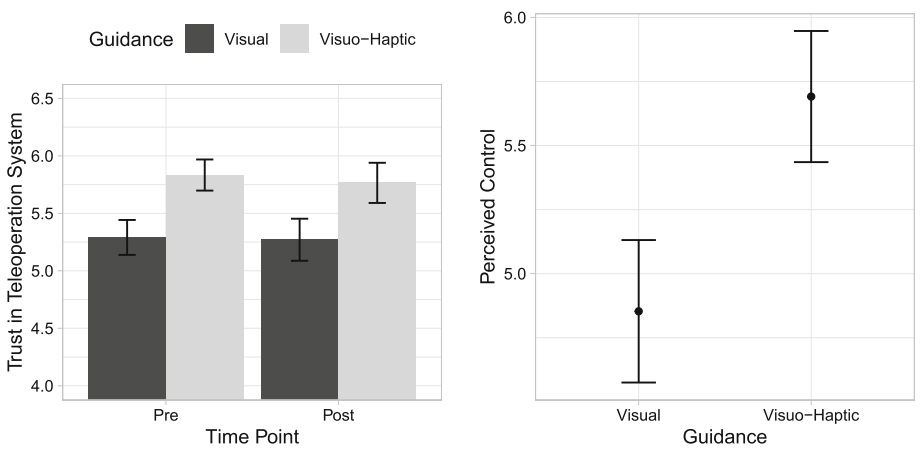
visuo-haptic guidance compared to visual guidance on participants' performance during such takeover situations, as well as their trust into teleoperation systems.

Our findings reveal important trade-offs of integrating haptic feedback into teleoperated autonomous systems. Visuo-haptic guidance significantly accelerated the detection of system failures compared to visual guidance alone. Furthermore, haptic feedback led to significantly higher trust in the teleoperation system and a stronger sense of perceived control. However, these benefits came at the cost of increased user workload, as participants reported higher physical demand and effort under visuo-haptic guidance, and haptic feedback did not improve task completion times or accuracy during the subsequent semi-autonomous control phase. This suggests that while haptic coupling keeps the user physically engaged and aware, it imposes a higher subjective effort without necessarily enhancing manual motor performance after the takeover. From an engineering perspective, these insights suggest potential avenues for designing future automated surgical cockpits. Because haptic guidance provides a physical connection during automated tasks, it could serve as a method to maintain user engagement and reduce the cognitive detachment associated with purely passive visual monitoring. However, to balance the observed physical workload, future surgical interfaces might explore adaptive or intermittent haptic frameworks rather than continuous tracking. For example, a system could remain haptically passive during routine automated sequences to prevent unnecessary user fatigue, while dynamically introducing active haptic cues or guidance forces only during high-uncertainty phases or critical takeover transitions. Furthermore, future studies should evaluate whether the observed advantage of visuo-haptic guidance over visual guidance persists in environments where the system explicitly signals uncertainty and failure, rather than requiring the operator to actively detect the event itself.

Crucially, our results regarding trust indicate that the higher trust observed with visuo-haptic guidance was established before the actual physical interaction. That is, trust for visuo-haptic guidance was higher than for visual guidance immediately after participants received the system description, and this difference persisted after task execution despite the occurrence of system failures. This pattern even extended to trust in autonomous systems in general, which increased after the experiment and thus after experiencing system failures. We attribute these findings to the predictability of the system's behavior, as the introductory material effectively set user expectations by explicitly stating the potential for stops. Consequently, the interaction itself confirmed the initial expectations rather than altering the user's trust, demonstrating that predictable system failures do not necessarily diminish trust if users are properly instructed.

There are specific limitations to this study that necessitate future research. First, we explored the role of haptic guidance specifically as a supplement to visual monitoring. We did not include a control condition without any guidance (i.e., where the operator does not monitor the system execution at all). Therefore, it remains an open question how performance and trust are affected when the operator is not even visually monitoring the system prior to a takeover request. Future work should investigate these lower levels of supervision to understand the baseline risks of distraction during autonomous phases. Second, as this study was conducted with non-expert students unfamiliar with the system and laparoscopic surgery, the results regarding workload and trust may differ for experienced surgeons. Future research should validate these findings with expert users on a set of clinically more relevant tasks, such as dissection, resection, needle insertion or suturing to ensure clinical transferability.

Author Contributions All authors contributed to the study conception, design, material preparation, and data collection. Analysis was performed by FP and KH. The first draft of the paper was written by FP,



(a) Trust in the Teleoperation System (b) Perceived Control

Fig. 6 Results on trust in the teleoperation system (**a**) and perceived control (**b**). Error bars indicate the standard error of the mean

KH and JK and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript

Funding Open Access funding enabled and organized by Projekt DEAL. This work was partially supported by the state of Bavaria through the "Bayerische Forschungsstiftung" (BFS, Grant "AZ-1592-23-ForNeRo").

Data Availability Data and codes are available from the authors upon reasonable request.

Declarations

Conflict of interest Frank Papenmeier, Katharina Hagmann, Lisa Blossey, Adrian S. Bauer, Andreea E. Potinteu, Alexander Kirst, Daniel Leidner, Markus Huff and Julian Klodmann declare to have no financial or non-financial interests to disclose.

Compliance with ethical standards This study was approved by the local ethics committee of the Leibniz-Institut für Wissensmedien (LEK 2024/024). Participants provided informed consent at the beginning of the experiment and agreed to the analysis of their data at the end of the experiment.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Haidegger T (2019) Autonomy for surgical robots: Concepts and paradigms. IEEE Trans Med Robot Bionics 1(2):65–76. https://doi.org/10.1109/TMRB.2019.2913282
2. Yang G-Z, Cambias J, Cleary K, Daimler E, Drake J, Dupont PE, Hata N, Kazanzides P, Martel S, Patel RV et al (2017) Medical robotics-Regulatory, ethical, and legal considerations for increasing levels of autonomy. American Assoc Adv Sci. https://doi.org/10.1126/scirobotics.aam8638
3. EuropeanParliament, Council: laying down harmonised rules on artificial intelligence and amending regulations (artificial intelligence act). Official Journal of the European Union (OJ L) **2024**(1689) (2024) (accessed: 2026-01-06)
4. Hagmann K, Hellings-Kuß A, Klodmann J (2023) Robot system and method for controlling a robot system. German Patent **DE102023124155A1**(A1) (accessed: 2026-01-06)
5. Caccianiga G, Mariani A, De Paratesi CG, Menciassi A, De Momi E (2021) Multi-sensory guidance and feedback for simulation-based training in robot assisted surgery: a preliminary comparison of visual, haptic, and visuo-haptic. IEEE Robot Autom Lett 6(2):3801–3808. https://doi.org/10.1109/LRA.2021.3063967
6. Zhang D, Tron R, Khurshid RP (2021) Haptic feedback improves human-robot agreement and user satisfaction in shared-autonomy teleoperation. 2021 IEEE International Conference on Robotics and Automation (ICRA). pp 3306–3312. https://doi.org/10.1109/ICRA48506.2021.9560991
7. Zhang B, Winter J, Varotto S, Happee R, Martens M (2019) Determinants of take-over time from automated driving: a meta-analysis of 129 studies. Transport Res F: Traffic Psychol Behav 64:285–307. https://doi.org/10.1016/j.trf.2019.04.020
8. Merat N, Jamson AH, Lai FCH, Daly M, Carsten OMJ (2014) Transition to manual: driver behaviour when resuming control from a highly automated vehicle. Transport Res F: Traffic Psychol Behav 27:274–282. https://doi.org/10.1016/j.trf.2014.09.005
9. Hoff KA, Bashir M (2015) Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 57(3):407–434. https://doi.org/10.1177/0018720814547570
10. Hancock PA, Billings DR, Schaefer KE, Chen JYC, Visser EJ, Parasuraman R (2011) A meta-analysis of factors affecting trust in human-robot interaction. Hum Factors 53(5):517–527. https://doi.org/10.1177/0018720811417254
11. Itoh M, Abe G, Tanaka K (1999) Trust in and use of automation: their dependence on occurrence patterns of malfunctions. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028), vol 3. pp 715–7203. https://doi.org/10.1109/ICSMC.1999.823316
12. Schaefer KE, Chen JYC, Szalma JL, Hancock PA (2016) A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. Hum Factors 58(3):377–400. https://doi.org/10.1177/0018720816634228
13. Seibold U, Kübler B, Bahls T, Haslinger R, Steidle F (2019) The dlr mirosurge surgical robotic demonstrator. In: The Encyclopedia of MEDICAL ROBOTICS: Volume 1 Minimally Invasive Surgical Robotics, pp. 111–142. World Scientific, N/A https://doi.org/10.1142/9789813232266_0005
14. Grespan L, Fiorini P, Colucci G (2019) The Route to Patient Safety in Robotic Surgery - Chap. 9 Surgical Robot Simulation and Control. Springer, N/A https://doi.org/10.1007/978-3-030-03020-9
15. Bowyer SA, Davies BL, Baena F (2014) Active constraints/virtual fixtures: a survey. IEEE Trans Robot 30(1):138–157. https://doi.org/10.1109/TRO.2013.2283410
16. Laubert T, Esnaashari H, Auerswald P, Höfer A, Thomaschewski M, Bruch H-P, Keck T, Benecke C (2018) Conception of the lübeck toolbox curriculum for basic minimally invasive surgery skills. Langenbecks Arch Surg 403(2):271–278. https://doi.org/10.1007/s00423-017-1642-1
17. Hwang M, Ichnowski J, Thananjeyan B, Seita D, Paradis S, Fer D, Low T, Goldberg K (2023) Automating surgical peg transfer: calibration with deep learning can exceed speed, accuracy, and consistency of humans. IEEE Trans Autom Sci Eng 20(2):909–922. https://doi.org/10.1109/TASE.2022.3171795
18. Malle BF, Ullman D (2021) A multidimensional conception and measure of human-robot trust. In: Nam CS, Lyons JB (eds.) Trust in Human-robot Interaction, pp. 3–25. Academic Press https://doi.org/10.1016/B978-0-12-819472-0.00001-0
19. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P. A. Hancock, N. Meshkati (eds.) Human Mental Workload, pp. 139–183. North-Holland, Amsterdam https://doi.org/10.1016/S0166-4115(08)62386-9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.